
Research Articles: Behavioral/Cognitive

Rapid visuomotor responses reflect value-based decisions

Timothy J. Carroll¹, Daniel McNamee², James N. Ingram^{2,3} and Daniel M. Wolpert^{2,3}

¹Centre for Sensorimotor Performance, School of Human Movement and Nutrition Sciences, The University of Queensland, Brisbane, 4072, Australia

²Computational and Biological Learning Laboratory, Department of Engineering, University of Cambridge, Cambridge, CB2 1PZ, UK.

³Zuckerman Mind Brain Behavior Institute, Department of Neuroscience, Columbia University, New York, United States.

<https://doi.org/10.1523/JNEUROSCI.1934-18.2019>

Received: 30 July 2018

Revised: 14 January 2019

Accepted: 15 January 2019

Published: 8 March 2019

Author contributions: T.J.C., D.M., J.N.I., and D.W. designed research; T.J.C. performed research; T.J.C., D.M., and D.W. analyzed data; T.J.C., D.M., J.N.I., and D.W. wrote the paper; J.N.I. contributed unpublished reagents/analytic tools.

Conflict of Interest: The authors declare no competing financial interests.

The authors thank Max Donelan for feedback on the manuscript and note funding support from the Australian Research Council (FT120100391), the Wellcome Trust and the Royal Society Noreen Murray Professorship in Neurobiology (to D.M.W.).

Correspondence should be directed to: Timothy Carroll, School of Human Movement and Nutrition Sciences, The University of Queensland, St Lucia QLD, 4072, Australia. Tel.: +61 7 3365 6380, Fax: +61 7 3365 6877, timothy.carroll@uq.edu.au

Cite as: J. Neurosci 2019; 10.1523/JNEUROSCI.1934-18.2019

Alerts: Sign up at www.jneurosci.org/alerts to receive customized email alerts when the fully formatted version of this article is published.

Accepted manuscripts are peer-reviewed but have not been through the copyediting, formatting, or proofreading process.

Copyright © 2019 Carroll et al.

This is an open-access article distributed under the terms of the Creative Commons Attribution 4.0 International license, which permits unrestricted use, distribution and reproduction in any medium provided that the original work is properly attributed.

1

2 Rapid visuomotor responses reflect value-based decisions

3 Abbreviated title: Visuomotor reflexes reflect value-based decisions

4 Timothy J. Carroll ¹, Daniel McNamee ², James N. Ingram ^{2,3}, Daniel M. Wolpert ^{2,3}

5

6 ¹ Centre for Sensorimotor Performance, School of Human Movement and Nutrition Sciences,
7 The University of Queensland, Brisbane, 4072, Australia8 ² Computational and Biological Learning Laboratory, Department of Engineering, University
9 of Cambridge, Cambridge, CB2 1PZ, UK.10 ³ Zuckerman Mind Brain Behavior Institute, Department of Neuroscience, Columbia
11 University, New York, United States.

12 Correspondence should be directed to:

13 Timothy Carroll
14 School of Human Movement and Nutrition Sciences
15 The University of Queensland
16 St Lucia QLD, 4072, Australia
17 Tel.: +61 7 3365 6380
18 Fax: +61 7 3365 6877
19 timothy.carroll@uq.edu.au
20

21 Number of pages = 32

22 Number of figures = 9

23 Number of tables, multimedia and 3D models = 0

24 Number of words for Abstract = 233

25 Number of words for Introduction = 647

26 Number of words for Discussion = 1609

27 Conflict of Interest

28 The authors declare no competing financial interests.

29 Acknowledgements30 The authors thank Max Donelan for feedback on the manuscript and note funding support
31 from the Australian Research Council (FT120100391), the Wellcome Trust and the Royal
32 Society Noreen Murray Professorship in Neurobiology (to D.M.W.).
33

34 **Abstract**

35 Cognitive decision making is known to be sensitive to the values of potential options, that is
36 the probability and size of rewards associated with different choices. Here, we examine
37 whether rapid motor responses to perturbations of visual feedback about movement, which
38 mediate *low-level* and involuntary feedback control loops, reflect computations associated
39 with *high-level* value-based decision making. In three experiments involving human
40 participants, we varied the value associated with different potential targets for reaching
41 movements by controlling the distributions of rewards across the targets (experiment 1), the
42 probability with which each target could be specified (experiment 2), or both (experiment 3).
43 We found that the size of rapid and involuntary feedback responses to movement
44 perturbations was strongly influenced by the relative value between targets. A statistical
45 model of relative value that includes a term for risk sensitivity provided the best fit to the
46 visuomotor response data, illustrating that feedback control policies are biased to favour
47 more frequent task success at the expense of the overall extrinsic reward accumulated
48 through movement. Importantly however, the regulation of rapid feedback responses was
49 associated with successful pursuit of high-value task outcomes. This implies that when we
50 move, the brain specifies a set of feedback control gains that enable low-level motor areas
51 not only to generate efficient and accurate movement, but also to rapidly and adaptively
52 respond to evolving sensory information in a manner consistent with value-based decision
53 making.

54 **Significance Statement**

55 Current theories of sensorimotor control suggest that, rather than selecting and planning the
56 details of movements in advance, the role of the brain is to set time-varying feedback gains
57 that continuously transform sensory information into motor commands by feedback control.
58 Here, we examine whether the fastest motor responses to perturbations of movement, which
59 mediate low-level and involuntary feedback control loops (i.e. reflexes), reflect computations
60 associated with high-level, value-based decision making. We find that rapid feedback
61 responses during reaching reflect the relative probabilities and rewards associated with
62 target options. This suggests that low-order components of the sensorimotor control
63 hierarchy, that generate rapid and automatic responses, can continuously evaluate evolving
64 sensory evidence and initiate responses according to the prospect of reward.

65

66 Introduction

67 Every decision that an animal makes must ultimately be implemented through movement.
 68 Therefore, rational decision making should take into account the properties and state of the
 69 motor system when weighing the desirability of options. Indeed, value-based decisions are
 70 sensitive to the physical costs of action (Croxxson et al., 2009; Skvortsova et al., 2014;
 71 Manohar et al., 2015; Klein-Flügge et al., 2016; Shadmehr et al., 2016), and decisions about
 72 which action to perform can be rapidly adapted if the state of the body unexpectedly
 73 changes (Nashed et al., 2012, 2014). There is both neurophysiological and behavioural
 74 evidence that the state of the motor system reflects decision variables prior to a final
 75 commitment to act (Platt and Glimcher, 1999; Song and Nakayama, 2008; Resulaj et al.,
 76 2009; Pastor-Bernier and Cisek, 2011; Selen et al., 2012). For example, the firing rates of
 77 neurons in the dorsal premotor cortex reflect the reward associated with a potential target in
 78 their receptive field relative to alternative targets (Pastor-Bernier and Cisek, 2011). Similarly,
 79 the gains of long latency stretch reflexes track evidence accumulation when a perceptual
 80 decision must be reported by a motor response to a change in limb position (Selen et al.,
 81 2012), and movements initiated under uncertainty are strongly biased by factors that are
 82 critical to evaluating the expected relative values of alternative choices, such as their
 83 associated rewards and costs (Seydell et al., 2008; Landy et al., 2012; Schütz et al., 2012).

84 Although these observations confirm a tight coupling between decision making and
 85 sensorimotor control, questions remain regarding the nature of the interactions between
 86 these processes. In particular, it is unclear to what extent decision variables modulate
 87 feedback control systems that are integral to effective movement. Our movements rely on
 88 flexible and hierarchical feedback control that effectively deals with noise and delays in
 89 sensory feedback by taking account of efference copy information. The selection of an
 90 “action” to be taken is therefore better conceived of as the specification of a feedback control
 91 policy that continuously transforms internal neural states and sensory inputs into motor
 92 outputs (Todorov and Jordan, 2002; Scott, 2004; Todorov, 2004). Such policies specify the
 93 gain of feedback loops at multiple levels of the sensorimotor hierarchy, such that all but the
 94 shortest latency spinal reflex arcs can be flexibly customised to the task context (Scott,
 95 2016). The assessment of rapid feedback responses to sudden changes in the state of the
 96 body or the environment therefore provides a window into the computations upon which
 97 feedback control policies are based. Here we use this approach to ask whether feedback
 98 control systems are influenced by the relative value of potential reach goals. If such decision
 99 variables are to effectively shape motor behaviour, their influence should be incorporated
 100 into feedback control policies and thus be observable in rapid feedback responses to sudden

perturbations. This would require that some aspects of value-based “decision making”, typically conceived of as a high-order cognitive computation, be implemented in the low-order components of the sensorimotor control hierarchy that generate rapid and automatic responses.

Here we examine whether rapid feedback responses to displacements of visual feedback of hand position are tuned to the relative values of alternative reach targets. Because the value of an option is the product of its contingent reward magnitude and probability, we first conducted separate experiments to determine whether rapid feedback responses are tuned to bring the hand closer to targets that carry greater rewards, and to targets presented more frequently than the alternatives. As both components of value modulated feedback responses, we conducted a third experiment involving different combinations of reward and probability, to test how well statistical models of reward, probability and relative value explain feedback response modulation. A relative value model with a term for risk sensitivity best fit the data, implying that low-level sensorimotor circuits can flexibly evaluate sensory information and rapidly tune motor responses according to the prospect of reward.

Materials and methods

Participants

Thirty-four self-reported, right-handed participants, without a reported neurological condition and with normal or corrected to normal vision, took part in one of three experiments (Experiment 1: 7 females, 5 males; ages ranged from 18–32 years; Experiment 2: 5 females, 7 males; ages ranged from 18–33 years; Experiment 3: 5 females, 5 males; ages ranged from 21–37 years). All participants were naïve to the purpose of the experiment. They received payment of between £20 and £35 with the amount depending on both experimental duration (1.5–2.5 hrs) and performance (Experiments 1 and 3; see below). The Cambridge Psychology Research Ethics Committee approved the experimental procedures which conformed to the Declaration of Helsinki. All subjects gave written informed consent.

Experimental Apparatus

Participants made reaching movements while grasping the handle of a robotic manipulandum (vBOT) that constrained motion of the hand to the horizontal plane. The vBOT is a custom built robotic device that measures the position of the handle and generates state-dependent forces at the handle endpoint (Howard et al., 2009). A six-axis

force transducer (ATI Nano 25; ATI Industrial Automation) measured the forces applied by the subject at the handle. Hand position was measured by optical encoders (58SA; Industrial Encoders Direct). Visual feedback was provided using a computer monitor (ASUS, VG278H, 120 Hz) mounted above the vBOT and was projected veridically to the subject via a mirror. Subjects were prevented from viewing their hand directly, and the virtual reality system was used to overlay images such as targets and hand cursor (0.5-cm-radius disk) in the plane of movement. Hand position and forces were sampled at 1000 Hz. The delay between position sampling and cursor display was measured with a photodiode as 24 ms at the centre of the screen, and all reported response times are corrected for the display latency.

Task Details

Each trial began with the participant's hand within a 0.55 cm radius home circle, which was aligned with the body midline. Participants made 25 cm reaching movement towards an array of three coloured boxes (each 6 x 5 cm; see Figure 1A). Participants were asked to fixate a small grey cross that was displayed in the central box for ~500 ms. The disappearance of the fixation cross was the go signal for the participant to initiate a reach toward the target array. If the hand speed did not exceed 10 cm.s^{-1} within 550 ms of the go signal, an error message was displayed ("Too late") and the trial was repeated. The low contrast of the dark grey cross on a black background made it difficult to perceive the go signal without fixation. To ensure the initial kinematics of movement were similar, participants were required to move the cursor through a small grey bar (2 cm wide x 1 cm long) located 11 cm from the home circle directly towards the center of the array. If the cursor did not pass through the intermediate grey bar, the screen turned red, all task information disappeared, and an error message was displayed ("Missed half-way target"). If the hand speed exceeded 60 cm.s^{-1} prior to reaching the intermediate grey bar, the red screen appeared and another error message was displayed ("Too fast"). Trials aborted due to kinematic errors at the intermediate marker were repeated immediately.

-- Insert Figure 1 about here --

On some trials, the cursor position was jumped laterally (left or right) by 3 cm when it reached the grey bar (the cursor was displayed on top of the bar). Critically, at this point, the final reach target was not yet specified. The final target was displayed as a yellow circle of 1.25 cm radius when the hand reached was 15.5 or 16 cm (see below) from the home position. The participant was required to bring the cursor within the target for at least 50 ms within 800 or 850 ms (see below) of movement initiation. Feedback of whether the trial was successful and the amount of reward (if relevant, see below) were provided if the target was

reached in time. If the target was not reached within the time limit, the cursor was extinguished and an error message was displayed (see below for details), but the trial was not repeated. In combination with the hand speed limit at the grey bar, the time limit served to constrain the peak speed of reaches to $\sim 50 \text{ cm.s}^{-1}$ (mean \pm SD, $50.8 \pm 2.5 \text{ cm.s}^{-1}$).

There were two types of trial for both the unperturbed and cursor jump conditions. In standard trials, no forces were applied by the robot to the handle during the reach. Thus, participants had to correct any cursor displacement by moving the hand laterally to reach the target (Fig. 1A). On channel trials, the robot constrained the handle to move along a straight path from the home position to the centre target (Fig. 1B) via a simulated stiff spring and damper (6000 N.m^{-1} , $100 \text{ N.m}^{-1}\text{s}$). On all channel trials in which the cursor jumped, the cursor was aligned to the selected target when it was 3 cm from the target. Therefore, when the central target was specified, the cursor was re-aligned with the hand. In these trials, we measured the lateral forces that participants exerted against the walls of the channel in response to the cursor jump as a measure of the reflex gain (Franklin and Wolpert, 2008; Franklin et al., 2016).

Experiment 1 - Reward Manipulation

The first experiment was designed to determine whether the magnitude of the earliest visuomotor response to a cursor jump is sensitive to the reward associated with potential targets, even before a target is specified. Participants completed blocks of 102 trials. Each block consisted of the following trials.

- a) 18 context trials: 6 standard unperturbed reaches to each of the three targets. These were performed first to illustrate the reward associated with each target in that block and were not analysed.
- b) 36 unperturbed trials: 12 reaches to each target which were standard except for 6 channel trials to the central target.
- c) 48 cursor jump trials: 24 left cursor jump trials with an equal number of left and central targets and 24 right cursor jump trials with an equal number of right and central targets. All trials were standard except for 6 channel trials to the central target for both cursor jump directions.

After the 18 context trials (a) that were performed at the start of each block, all remaining trials (b & c) were performed in a random order. Participants completed 4 familiarisation and 6 experimental blocks. In the familiarisation blocks, task success was signalled only by the display of the text "Hit" or "Miss" immediately after each trial, and the colour of the boxes surrounding the target array was varied randomly. The familiarisation blocks allowed

202 participants to learn the task and were also used to identify the appropriate time constraint
 203 for target acquisition (800 or 850 ms) and the hand position at which the target was revealed
 204 (15.5 or 16 cm). The goal was to make the task difficult for each participant, as pilot data
 205 showed considerable inter-individual differences in performance, and that reward-related
 206 effects were minimal if subjects could reach the target on every trial. We chose a setting for
 207 movement duration and position at which target was specified for the remaining blocks to
 208 obtain a hit rate of 70-85% (across all three experiments, mean \pm SD hit rate on standard
 209 trials was $73.8 \pm 9.8\%$).

210 In the experimental blocks, rewards associated with potential targets were signalled by the
 211 colors of the three boxes. Targets that appeared inside boxes of one color (red or blue,
 212 counterbalanced across subjects) were associated with high reward (£10), whereas targets
 213 that appeared inside a box of the other colour were associated with low reward (50p). In six
 214 alternating blocks, either the two outer targets were worth £10 and the centre target worth
 215 50p, or vice versa (Figure 1C: reward ratios of 1/20 or 20). If a high reward target was hit, an
 216 image of a £10 note was displayed on the screen and a pleasant “ding” sound was played. If
 217 a low reward target was hit, an image of a 50p coin was displayed and no sound was played.
 218 If the target was missed, the text “Too late – no bonus” was displayed. Subjects were
 219 informed that one trial in which each reward type had been available would be selected at
 220 random at the conclusion of the experiment. If the trial of that type was hit, then they would
 221 receive the associated reward (£10 or 50p), whereas if the trial was missed, no reward
 222 would be given. All participants were given £20 to compensate them for their time, so the
 223 final payment made to each person could be £20, £20.50, £30 or £30.50.

224 A final 11th block was included to measure the size and latency of the rapid visuomotor
 225 response to cursor jumps in the absence of target uncertainty. The block involved reaches to
 226 a single, central target that was visible throughout each trial. The first 12 trials were standard
 227 unperturbed trials and were not analysed, and the remaining 90 trials comprised of 30
 228 standard unperturbed trials, 12 standard trials each with left and right cursor jumps, 12
 229 unperturbed channel trials, and 12 channel trials each with left and right cursor jumps. The
 230 reward associated with all trials was 50p.

231 **Experiment 2 - Probability Manipulation**

232 The second experiment was designed to determine whether the magnitude of the earliest
 233 visuomotor response to a cursor jump is sensitive to the probability with which potential
 234 targets appear. In this experiment, no financial rewards were contingent upon task
 235 performance. Participants first completed 4 familiarisation blocks that were identical to those
 236 in Experiment 1, followed by 6 blocks in which either the centre target or the two outer

237 targets appeared more frequently, and one final block that corresponded to the last block of
 238 Experiment 1, in which a single target was visible throughout the reach. As above, these
 239 familiarisation trials were used to keep target hit rates at ~70-85% in the remainder of the
 240 experiment.

241 In the six blocks with probability manipulation, the probability of each target being presented
 242 was signalled by the color of the box surrounding each potential target location. Targets
 243 inside boxes of one color (red or blue, counterbalanced across subjects) were presented
 244 twice as often as those inside boxes of the other colour. That is, the probability of
 245 left:centre:right targets was either 0.4: 0.2: 0.4 or 0.25: 0.50: 0.25. When a target was hit, the
 246 text “hit” was displayed on the screen and a pleasant “ding” sound was played. If the target
 247 was missed, the text “Miss” was displayed and no sound was played. All participants were
 248 given £20 if the experiment lasted up to 2 hrs, plus £2.50 for any additional periods of 15
 249 min.

250 Each block consisted of:

251 a) 12 unperturbed context trials with targets presented exclusively within coloured boxes
 252 defining the more likely location (in alternating blocks: either in the centre or at the two outer
 253 locations). These trials were performed first and were included to illustrate the location of the
 254 more probable targets, and were not analysed.

255 b) The number of remaining trials in each block depended on whether the centre or outer
 256 targets were presented more frequently in that block. In all cases, half the trials to the centre
 257 target were channel trials.

258 If the centre target was more probable, the remaining 100 trials included 40 unperturbed
 259 trials (target numbers 10:20:10) and 60 cursor jump trials with (10:20:0) for left and (0:20:10)
 260 for right cursor jumps.

261 If the outer targets were more probable, the remaining 110 trials included 50 unperturbed
 262 trials (20:10:20) and 60 cursor jump trials with (20:10:0) for left and (0:10:20) for right cursor
 263 jumps.

264 **Experiment 3 - Value Manipulation**

265 The third experiment included manipulations of both reward and probability, in order to
 266 determine whether the magnitude of the earliest visuomotor response to a cursor jump
 267 scales with the relative value of potential targets. The first block of 102 trials was identical to
 268 the single trials blocks of Experiments 1 and 2.

269 Participants then completed two familiarisation blocks of 120 standard trials consisting of 24
 270 trials each of (1) unperturbed, (2) left and (3) right cursor jumps to the centre target, and (4)
 271 left and (5) right cursor jumps to the left and right targets, respectively. As above, these
 272 familiarisation trials were used to keep target hit rates at ~70-85% in the remainder of the
 273 experiment. A key difference in this design compared to the first two experiments is that we
 274 included channels on trials in which the outer target was specified after a cursor jump. On
 275 these trials, the cursor was jumped again to align with the specified (outer) target when it
 276 was 3 cm from the target.

277 Participants performed eight experimental blocks in which the relative probability and relative
 278 rewards were varied. After a cursor jump, the relative probability (P_O/P_C) of the outer versus
 279 centre target being specified was either 0.5, 1 or 2, and the relative reward for the outer
 280 versus centre target (R_O/R_C) was either 0.5, 1 or 2. We examined all combinations (Fig.1C)
 281 of relative reward and probability in different blocks (except where both were 1). Blocks
 282 were performed in pairs with $R_O/R_C=r$ and $P_O/P_C=p$ for one block followed by $R_O/R_C=1/r$ and
 283 $P_O/P_C=1/p$ for the other. The pairs were performed in a pseudorandom order.

284 The first 18-20 trials of each block were standard trials that served to illustrate the reward
 285 and probability characteristics of the block. These initial trials were not analysed. The
 286 remaining 118 trials involved 108 cursor jump trials and 10 unperturbed trials, randomly
 287 intermixed. The centre target was cued on all the unperturbed trials; 5 were channel trials
 288 and 5 were standard trials. The 108 cursor jump trials comprised 72 standard trials and 36
 289 channel trials. Half of these involved a left cursor jump, and half involved a right cursor jump.
 290 The number of trials in which the centre and outer targets were cued on these cursor jump
 291 trials is defined by the relative probabilities shown in Figure 1C. Either the centre target was
 292 cued twice as often as the outer target for each direction of cursor jump ($P_O/P_C = 0.5$), cued
 293 half as often as the outer target for each cursor jump direction ($P_O/P_C = 2$), or cued the same
 294 number of times for each cursor jump direction ($P_O/P_C = 1$). Due to a coding error, one trial in
 295 each block of trials was assigned to an incorrect condition (with respect to the probabilities
 296 defined above). This resulted in an average deviation from the intended probability ratios of
 297 $0.19\% \pm 0.56$ (group mean \pm SD).

298 As in Experiment 1, rewards associated with potential targets were signalled by the colors of
 299 the three boxes surrounding the potential target array. Boxes of one color (red or blue,
 300 counterbalanced across subjects) were associated with high reward (£10), and boxes of the
 301 other colour were associated with low reward (£5). If a high reward target was hit, a large
 302 image of a chest of gold and the text “£10” were displayed on the screen and a pleasant
 303 “ding” sound was played. If a low reward target was hit, a (50%) smaller image of chest of

gold was displayed, the text “£5” was displayed, and a brief, medium-pitched tone was played. If the target was missed, the text “Too late – no bonus” was displayed. Subjects were informed that one trial in which each reward type had been available would be selected at random at the conclusion of the experiment. If the trial of each type had been hit, then they would receive the associated reward (£10 or £5), whereas if the trial was missed, no reward would be given. All participants were given £20 to compensate them for their time, plus £2.50 for any additional periods of 15 min, so the final payment made to each person could be £20, £25, £30, or £35 in addition to extra time payments.

Experimental Design and Statistical Analysis

The velocity and force (on channel trials) from each trial were low-pass filtered at 50 Hz, and the filtered velocity differentiated to provide endpoint acceleration. The average lateral force time series exerted by each subject during unperturbed channel trials was subtracted from forces exerted on each channel trial in which there was a cursor jump. Similarly, the average acceleration time series recorded during unperturbed standard trials were subtracted from standard cursor jump trials. The primary measure of the short latency response to a visuomotor displacement was the average (subtracted) lateral force exerted against the walls of the force channel between 170 and 220 ms after the cursor was displaced. We also calculated the average lateral acceleration of the hand in the same time window after cursor displacement in standard trials.

The decision to use of an analysis window of 170-220 ms was taken in order to make our analysis comparable to the majority of recent papers that used the cursor jump paradigm (Dimitriou et al., 2013; Franklin et al., 2014, 2016; Gallivan et al., 2016)). Note that voluntary responses to cursor jumps occur ~320 ms after the perturbation (Franklin and Wolpert, 2008), and voluntary responses to target jumps occur ~220 ms after the perturbation (Day and Lyon, 2000). However, it is also of interest to determine the onset time at which any effect of target value becomes evident. To this end, we used signal detection theory on a subject by subject basis to identify the time at which the force responses to cursor jumps first diverged between blocks in which the outer and centre targets were more valuable. This method was based on the one used by (Weiler et al., 2015) with two modifications. We generated an ROC curve for every 1-ms sample and calculated the area under the ROC (aROC) curves for the ability to distinguish between the responses for which outer and centre targets were more valuable. As we are interested in the time point at which this difference emerges in the force responses, we examine the time point where the information begins to deviate from chance. To do this, we excluded aROC after the point when the

aROC exceeded 0.62 for three consecutive samples. This is reduced from the values of 0.75 used in (Weiler et al., 2015) as the onset of the target (which occurred after than hand had moved a further 4.5 cm from the cursor jump location, that is around 100 ms later) affected the movement so that some curves did not reach a value of 0.75. We then fit a dog-leg to the aROC data (flat line at aROC of 0.5 followed by a linear component). In addition, given the lower aROC criterion, to be conservative in our estimate of the divergence time, we took the later of two times as the onset time of the response: 1) the end of the flat portion of the fit, and 2) the last local minimum in the aROC curve. This method provides a sensitive and conservative measure of the first time at which the visuomotor responses are modulated by expected value. We compared this time with that of the first force divergence between cursor jump and unperturbed trials in the single target experiment block, to establish whether modulation of value-based response magnitude is evident from the earliest component of the visuomotor response.

In experiments 1 and 2, the central target was selected on all channel trials, and we only analysed corresponding standard trials in which the centre target was presented. Group effects were analysed with 2-way repeated measures analysis of variance (3 block types [central target more rewarded / probable, outer targets more rewarded / probable, single target control trial] x 2 cursor jump directions [left, right]). Effect sizes are reported as partial Eta-squared (η^2) statistics for relevant main and interaction effects. Greenhouse-Geisser corrections were applied to the degrees of freedom where violations of the assumption of sphericity were detected, and Holm-Bonferroni post-hoc contrasts were used to assess pairwise differences between means. Statistica version 13.2 (Dell Software, Round Rock, Texas) was used for these tests.

Experiment 3 differed in that some reaches were channel trials even though an outer target was presented. To confirm that the eventual location of the target did not affect the early force response to cursor jumps within our analysis window of 170-220 ms, we again used signal detection theory to provide a sensitive measure of the first time at which the force response to central and outer targets began to diverge after a cursor jump. Note that the target was not revealed until the hand had moved a further 4.5-5 cm (~100 ms) beyond the point at which the cursor jump was applied. We used the ROC approach described above to distinguish between trials in which outer and centre targets were presented, and found that force response diverged after ~230 ms (mean \pm SD initiation time = 233 ± 15 ms for left cursor jumps; 228 ± 13 ms for right cursor jumps).

372 **Hierarchical Bayesian Modeling**

373 To test the extent to which early responses to visual perturbation vary systematically with the
 374 reward, probability and value of potential targets in Experiment 3, we performed hierarchical
 375 Bayesian modelling using custom Python scripts. Specifically, we fitted a hierarchical
 376 Bayesian model of group-level and subject-specific parameters in our models of response
 377 gains (Kruschke, 2010). These parameters were the slopes and intercepts linearly relating
 378 log value, probability, or reward to gains as well as the weighting parameter in the risk-
 379 sensitive model. Analyzing the data in this manner facilitates the identification of
 380 heterogeneous probability, reward, and risk attitudes across the population (as captured by
 381 the subject-level parameters) while making parameter estimation more robust to noise in
 382 response gains by partially pooling variance across subjects (due to the use of a prior over
 383 group-level variance). At the group-level, we specified weak priors for the mean (Gaussian,
 384 mean = 0, standard deviation = 10000) and variance (half Cauchy, mean = 0, scale = 5) for
 385 all parameters. All priors were selected independent of the data based on previously
 386 established defaults used in hierarchical modeling analysis. At the subject-level, parameters
 387 for each subject were modeled separately using Gaussian (probability, reward, and value
 388 parameters) or bounded Gaussian (risk weighting parameter) distributions with means and
 389 variances drawn from the estimated group-level distributions. Error variance prior was
 390 modeled using a half Cauchy distribution (scale = 5). Parameters were fitted by maximum a
 391 posteriori estimation. The posterior was approximately inferred using Markov Chain Monte
 392 Carlo sampling based on the No-U-Turn algorithm (Salvatier et al., 2016). The sampler was
 393 tuned using 1000 samples, 30000 samples were then drawn. After discarding the first 5000
 394 samples which, due to non-stationarity, are relatively unlikely to be representative of the
 395 posterior, the remaining 25000 were used for parameter estimation. Chain convergence was
 396 established based on the R-hat diagnostic values (which were very close to 1) and visual
 397 inspection of the sample traces.

398

399 **Code Accessibility**

400 All code used for data analysis will be provided upon written request to the corresponding
 401 author.

402

403

Results

Experiment 1 - Reward Manipulation

The first experiment was designed to determine whether the magnitude of rapid feedback responses to displacements of visual feedback of the hand location is sensitive to the rewards associated with equally probable reach targets. Participants grasped the handle of a robotic interface, and received visual feedback of the position of their hand via a cursor that was overlaid into the plane of movement. They were required to initiate reaches toward the centre of a three-target array (Figure 1A), and the target was specified from the three alternatives late in the movement. The three targets were associated with different rewards; either the central target was associated with 20 times (£10 versus 50p) greater reward than the two outer targets, or vice versa. On some trials (4 out of every 7 trials), the cursor was displaced (i.e. jumped) 3 cm to the left (example in Figure 1A) or right of the hand location just prior to the midpoint of the movement (and 102.9 ± 9.3 ms, mean \pm SD, before the target was specified). The experiment was designed (Figure 1C) so that on these trials the centre target and the outer target, towards which the cursor jumped, had equal probability of being specified (and the opposite outer target was never specified). On trials in which the cursor did not jump all three targets were equally likely. We also ran a control condition in which only a single central target was presented.

On standard trials the hand was unconstrained by the robotic interface (Figure 1A) and participants could correct for cursor jumps. To assess the strength of the visuomotor response to such a perturbation, a force channel was used on a subset of the trials (channel trials; Fig 1B) in which the centre target was cued, to constrain the hand to a straight-line path to the target. The strength of the response was then assessed as the mean lateral force exerted into the wall of the channel from 170-220 ms after the cursor jump was applied. A rapid, involuntary response is known to start at around 150 ms for single target trials, and the voluntary response occurs at approximately 320 ms (Franklin and Wolpert, 2008). Therefore, our measure captures the earliest response to visual error that is automatic and not under conscious control.

Figure 2 shows the lateral forces that participants exerted in the force channel just after the cursor jumped either left or right. The data for right and left cursor jumps have been combined by reversing the sign of the force for rightward cursor jumps, and the traces have been truncated at 250 ms to isolate the early latency response to the cursor jump. Figure 1A shows force traces from each individual trial in the three conditions for an example participant, and illustrates that visuomotor responses to the cursor jump began at around

125 ms in the control condition in which only the centre target was rewarded (ROC analysis showed that perturbed force responses first deviated from force responses on no perturbation trials at 126 ± 11 ms, mean \pm SD). Notably, some early responses were in the same direction as the cursor jump, especially in trials in which the outer targets were more rewarded. The effects of reward can be seen more clearly when average force traces for each participant are plotted in Figure 1B, and when the group averages (\pm SE) are plotted in Figure 1C. For the control condition (1 block of 102 trials), in which only the central target was displayed (grey line), the trace shows a typical strong visuomotor response that is in the opposite direction to the cursor jump (positive force), thereby trying to return the hand to the central target. For the blocks in which all three targets were possible (6 blocks of 102 trials, alternating between blocks in which the centre and outer targets were rewarded more), the cursor jump moved the cursor closer to one of the outer targets and away from the centre target. In such cases (red and blue), the average responses were weaker than in the single target condition. This reflects the variability in response amplitude and direction observed across trials and participants when there was more than one potential target (blue and red traces in Figures 1A and 1B). This may be because the ultimate target was uncertain, and if the final target happened to be the outer target towards which the cursor jumped, there would be less need to respond to the jump. Splitting the responses by whether the centre (red trace) or outer (blue trace) targets were rewarded more shows that the responses are biased towards the more rewarded target.

458

459 -- Insert Figure 2 about here --

460

Figure 2D shows the group results for the average lateral force exerted from 170-220 ms after the cursor jump. As reflected in the force-time plots, force responses that would tend to correct the cursor jump and bring the cursor back toward the centre target were greatest for the single target control condition. Strikingly, however, when the outer targets were more highly rewarded, the group average response was in the same direction as the cursor jump towards the outer target, that is a reversal of the normal reflex response. The sign of the response was not uniform across participants, and not statistically different from zero at group level (single sample t-test, $t_{11} = -2.2$, $P = 0.054$). However responses were significantly biased in the direction of the cursor jump for 7 of the 12 participants ($P < 0.05$; single sample t-tests against 0 for each subject, uncorrected for multiple comparisons), which shows that short latency responses to visuomotor error signal were reversed in some people. Moreover, the trend for corrective responses to be largest in single target trials, followed by conditions

in which the centre and then the outer targets were more highly rewarded, was highly consistent across subjects. A repeated measures anova (3 target conditions x 2 cursor jump directions) gave a significant effect of condition ($F_{1,3,14,2} = 29.4$, $P < 10^{-4}$, $\eta^2 = 0.73$). There were no other significant main or interaction effects (both $P > 0.20$). Holm-Bonferroni post hoc contrasts showed that response size was significantly greater for the control condition than either of the asymmetric reward conditions (both $P < 10^{-3}$), and that the corrective force towards the centre target was greater when the centre target was more highly rewarded than the outer targets ($P < 0.01$).

An ROC analysis was used to identify the first time after the perturbation of visual feedback that the force responses differed for trials performed when the outer versus the centre targets were more rewarded. The vertical dotted green line (mean with shaded \pm SE) shows that the outer and inner rewarded trials began to diverge at 163 ms, which was significantly later than the onset of the response determined in single target trials (paired t-test, $t_{11} = -3.3$, $P = 0.007$). Thus, the effect of reward on the rapid feedback response was not reliably apparent until ~ 35 ms after response onset. This timing discrepancy raises the possibility that there are multiple components to fast visuomotor responses to cursor jumps (i.e. perhaps analogous to the multiple components of long-latency stretch reflexes), and that the earliest response component is not affected by reward. To our knowledge, however, in no previous work has rapid visuomotor responses been dissociated into multiple components. Moreover, the early components of the responses are smallest in magnitude and therefore more subject to noise. We therefore interpret these timing effects cautiously at this stage, and leave the resolution of this intriguing question for future work. Critically, even if there are multiple components to rapid visuomotor responses, the effect of reward was apparent within our analysis window of 170-220 ms. This is well before the onset of voluntary responses to cursor jumps (~ 320 ms, (Franklin and Wolpert, 2008)), and lies within the standard epoch typically used to measure automatic and involuntary feedback responses. An analysis of the lateral acceleration on standard trials and in which either the outer target or the central target could have been cued showed a very similar separation of responses (see Figure 2E & F). To simplify comparison with the channel trials, we only analysed the trials in which the centre target was eventually cued. Again, there was a significant effect of condition ($F_{1,2,12,9} = 38.3$, $P < 10^{-4}$, $\eta^2 = 0.78$), and no other significant main or interaction effects (both $P > 0.4$). Holm-Bonferroni post hoc contrasts showed that response size was significantly greater for the control condition than either of the asymmetric reward conditions (both $P < 10^{-4}$), and that the corrective acceleration towards the centre target was greater when the centre target was more highly rewarded than the outer targets ($P < 0.01$). These data confirm that the tendency for rapid visuomotor responses to be biased toward high

509 value targets is not specific to movements made in a force channel, but are also apparent
510 when a change in the the limb trajectory was required to correct the visual error.

511 To determine whether early, involuntary responses to a cursor jump were important for task
512 performance, we also compared hit-rates for targets that were associated with high and low
513 rewards. We focussed particularly on standard, unconstrained trials in which there was a
514 cursor jump, when the participants were required to correct the cursor trajectory in order to
515 hit a target. When pooled across centre and outer targets, the hit rate was significantly
516 greater for trials in which the cued target was associated with the high ($77 \pm 10\%$, mean \pm
517 SD), rather than low reward ($69 \pm 12\%$; paired t-test, $t_{11} = 2.9$, $P = 0.015$). Thus, the bias in
518 rapid feedback response towards more highly rewarded targets was associated with more
519 successful acquisition of high reward targets. This suggests that feedback control policies
520 can be rapidly (i.e. in less than 100 trials over a few minutes) modified to increase the value
521 of motor outcomes, in the absence of changes in the physical characteristics of the task.

522 **Experiment 2 - Probability Manipulation**

523 In the second experiment, there was no monetary reward associated with the targets, but we
524 varied the relative probability of the central versus outer targets in separate blocks. The
525 experiment was designed (Figure 1C) so that on cursor jump trials within a block, the
526 probability P_C of the centre target being specified was either half or twice the probability P_O
527 of the outer target (towards which the cursor jumped) being specified. Similarly, on trials in
528 which the cursor did not jump, each outer target had either twice or half the probability of
529 being presented compared to the centre target. There were three blocks of 100 trials with
530 $P_O < P_C$, which were alternated with three blocks of 110 trials with $P_C < P_O$.

531 Figures 3A,B,C show the force traces from these two conditions and from the single target
532 condition. The results are strikingly similar to those observed for reward manipulations.
533 Again, a robust early response was seen in the single target condition (grey), and the
534 responses were smaller when there was target uncertainty (red and blue). Figure 3C also
535 shows that the onset time of the response in the single target condition, determined by ROC
536 analysis, was 125 ms (the dashed vertical black line with grey shading shows mean \pm SE).
537 When the outer targets were more probable, the cursor jump led to an average response
538 that would have brought the hand closer to the outer target. In contrast, when the centre
539 target was more probable the response was weaker. The ROC analysis used to identify the
540 first time at which the force responses differed for trials performed when the outer versus the
541 centre targets were more probable showed that trials obtained when the outer and inner
542 targets were more probable trials began to diverge at 151 ms (vertical dashed green line
543 with shading shows mean \pm SE). As was the case for reward asymmetry revealed in

experiment 1, the effect of probability on visuomotor response magnitude was significantly delayed from the onset of the response in single target trials (paired t-test, $t_{11} = -2.5$, $P = 0.03$).

Figure 3B shows the group results for the average lateral force exerted from 170-220 ms after the cursor jump. As was the case for more highly rewarded targets, when the outer targets had a higher probability of being selected, the group average response was in the same direction as the cursor jump; towards the outer target. Here, the effect was statistically different from zero at group level ($t_{11} = -2.3$, $P = 0.04$), and responses were significantly biased in the direction of the cursor jump for 5 of the 12 participants ($P < 0.05$; single sample t-tests against 0 for each subject). Once again, the trend for corrective responses to be largest in single target trials, followed by conditions in which the centre and then the outer targets were more probable, was highly consistent across subjects (Figure 3B). A repeated measures anova (3 target conditions x 2 cursor jump directions) gave a significant effect of condition ($F_{1,3,14.2} = 38.7$, $P < 10^{-5}$, $\eta^2 = 0.78$). There were no other significant main or interaction effects (both $P > 0.2$). Holm-Bonferroni post hoc contrasts showed that response size was significantly greater for the control condition than either of the asymmetric probability conditions (both $P < 10^{-4}$), and that the corrective force towards the centre target was greater when the centre target was more probable than the outer targets ($P < 0.01$).

An analysis of the lateral acceleration on standard trials showed a similar separation of responses (see Figure 3C & D). Again, there was significant effect of condition ($F_{1,5,16.3} = 42.0$, $P < 10^{-6}$, $\eta^2 = 0.79$), and no significant main effect of cursor jump direction ($F_{1,11} = 0.17$, $P = 0.6$). However, there was a significant interaction effect between the probability condition and cursor jump direction, such that greater response sizes were apparent for leftward than rightward cursor jumps ($F_{1,2,13.7} = 5.7$, $P = 0.03$, $\eta^2 = 0.34$). Holm-Bonferroni post hoc contrasts showed that response size was significantly greater for the control condition than either of the asymmetric probability conditions irrespective of cursor jump direction (all $P < 10^{-4}$). Despite the significant interaction effect, the corrective acceleration towards the centre target was significantly greater when it was more probable than the outer targets for both leftward cursor jumps ($P < 10^{-4}$) and rightward cursor jumps ($P = 0.02$). This subtle lateral asymmetry presumably reflects biomechanical factors associated with the impedance of the limb, since the corresponding force channel responses were symmetric.

As was the case for the target reward manipulation, hit-rates for targets that were more probable ($82 \pm 9\%$, mean \pm SD) were significantly greater than for targets that were less probable ($73 \pm 14\%$; paired t-test, $t_{11} = 5.4$, $P = 2.2 \times 10^{-4}$). Thus, rapid feedback response size was biased towards targets that were either more probable or more highly rewarded,

and in both cases modulation of automatic visuomotor responses was associated with more successful acquisition of high-value targets. Since the value of an option is determined both by reward magnitude and how often the reward is available, we designed a third experiment to determine whether these results based on independent reward and probability manipulations can be unified within the framework of value-based decision making.

-- Insert Figure 3 about here --

Experiment 3 - Value Manipulation

In a third experiment we varied both the relative reward and probability of specifying the centre or outer target. The relative reward for the outer versus central target (R_o/R_c) was either 0.5, 1 or 2 in separate blocks. After a cursor jump, the relative probability (P_o/P_c also fixed within a block) of the outer versus centre target being specified was either 0.5, 1 or 2 (Fig 1C). We examined all combinations of relative reward and probability in different blocks (except where both were 1; hence 8 blocks). These combinations led to 5 different relative values (i.e. the product of relative reward and probability) between the centre and outer targets. In contrast to the first two experiments, we included channels trials in which the outer target was specified after a cursor jump. On these trials, the cursor jumped to the specified (outer) target at the end of the movement (i.e. at 22 cm, long after the response had been measured). We used ROC analysis (see methods) to compare these trials, split by whether the centre or the outer target was specified. This showed that the force response diverged on average 233 and 228 ms (for left and right targets) after the perturbation (Figure 4). As this was outside our window for measuring the visuomotor response, we combined these trials in our analysis.

-- Insert Figure 4 about here --

Figure 5 shows the average force traces from all eight conditions split in separate panels for the three reward ratios (Fig. 5A) and for the three target probability ratios (Fig. 5B). In general, the results show a tendency for greater corrective responses toward the centre target for conditions in which the centre target was more probable (Fig 5A). The trend is less clear-cut for comparisons between conditions with different relative rewards, presumably due partly to the fact that the relative reward ratios were only 0.5, 1, and 2 in Experiment 3, rather than 0.05 and 20 in experiment 1. The results for acceleration responses made in

611 standard, non-channel trials are very similar to the force responses in channel trials (see Fig.
612 6).

613 -- Insert Figures 5 and 6 about here --
614

615 We first examined the visuomotor response as a function of only the reward ratio or only the
616 target probability ratio. The aim was to compare the modulatory effect of reward and
617 probability on the visuomotor response with those observed in experiments 1 and 2.
618 Consistent with experiments 1 and 2, regression analyses across conditions confirmed that
619 the response size depended significantly on both variables separately ($P < 10^{-4}$ & $P = 0.009$
620 respectively). Since the ratios of the experimental parameters varied nonlinearly across
621 conditions in this experiment, we chose to use a log scale on which they are spaced linearly,
622 thus facilitating model fitting and interpretation. Furthermore, neural recordings indicate that
623 probabilistic information about which of two potential visual targets is correct on a given trial
624 is encoded by neurons in a log probability ratio (Yang and Shadlen, 2007).

625 -- Insert Figure 7 about here --
626

627 Given that both probability and reward significantly affected the visuomotor response gain
628 individually, we examined how the gain modulation depended on the combination of the two
629 in the form of expected value (the product of reward and probability). A rational agent,
630 seeking to maximise cumulative reward, might modify the response gain according to log
631 relative value:

$$632 \quad \log RV = \log \frac{P_o R_o}{P_c R_c} = \log \frac{P_o}{P_c} + \log \frac{R_o}{R_c} \quad (1)$$

633 However, when comparing option values which vary based on reward and probability,
634 human decision-making tends not to be objectively rational according to expected value
635 theory (Kahneman, 2003). In fact, humans can be risk-seeking or risk-averse in
636 sensorimotor decision tasks (Nagengast et al., 2010, 2011a, 2011b; Braun et al., 2011). That
637 is, if participants are risk-seeking, they may have a higher gain for the more rewarding target
638 even if the probability of its specification is low. In contrast, risk-averse participants may
639 always have a higher gain towards the more probable target even if it has lower reward. In
640 order to model possible heterogeneity in risk sensitivity across subjects, a weighting
641 parameter w was introduced which was restricted to take values between 0 and 1:

$$\log RV = 2 \left[w \log \frac{P_o}{P_c} + (1 - w) \log \frac{R_o}{R_c} \right] \quad (2)$$

A value of $w=0.5$ corresponds to rational risk neutral behaviour (as in Eqn. 1), while a value $w<0.5$ implies that a subject is risk seeking and is motivated more by the reward ratio. Finally, a value $w>0.5$ implies risk aversion. It is also possible to interpret the w parameter as changing the relative value by an “intrinsic” reward (i.e. a reward associated with successfully reaching the target independent of the explicit experimenter-controlled reward; see Discussion).

We fit both the risk-neutral (Fig. 6C) and risk-sensitive (Fig. 6D) models. Given the sensitivity of model fitting to noise in the response gains, and the possible heterogeneity of risk attitudes across subjects, we fit the model parameters at both the individual subject-level and at the group-level via approximate Bayesian inference. This approach flexibly optimizes partial pooling across subjects at the group-level, while adapting to subject-specific parameter differences (see Methods). We measured the quality of model fit using the Watanabe-Akaike Information Criterion (WAIC) as it is sensitive to the full posterior over the fitted parameters (as opposed to a point estimate as in other criterion measures) and has been shown to be more precise in estimating out-of-sample predictive accuracy (Vehtari et al., 2017). Model comparison showed that the risk-sensitive relative value model (WAIC=7129) provided a better fit to the response gain data than any of the other models tested (risk-neutral relative value model WAIC=7136, relative probability model WAIC=7133, relative reward model WAIC=7178). Estimating pairwise Bayes factors from WAIC scores results in factors of $K_{\text{val}}=33$, $K_{\text{prob}}=7$, $K_{\text{rew}}>150$ when comparing the risk-neutral relative value, probability, and reward, models respectively to the risk-sensitive relative value model. These Bayes factors indicate strong, positive, and very strong evidence in favour of the risk-sensitive relative value model respectively (Kass and Raftery, 1995). With respect to the risk-sensitive relative value model, we examined risk sensitivity in visuomotor gain modulation by testing whether the weighting parameter in Eqn 2 was significantly different from 0.5 (a weighting parameter of 0.5 would be equivalent to risk-neutral gain modulation as in Eqn. 1). The weighting parameter w was estimated to be significantly greater than 0.5 (mean 0.68, 95% credible interval [0.52,0.86]) at the group level, indicating significant risk-aversion (or a non-zero contribution to value from a putative intrinsic reward associated with task success). Figure 6 shows the posterior predictive fits (i.e. the predicted visuomotor gains and estimation uncertainty) of the risk-neutral (Fig. 6C) and risk-sensitive (Fig. 6D) models with 50% and 95% credible regions in addition to probability (Fig. 6A) and reward (Fig. 6B) modulation models for comparison. The models were used to compute the mean decision parameter (e.g. risk-modulated log relative value in the case of panel D) and

677 response gain for each subject and unique experimental condition. The mean across
 678 participants for each condition are plotted on top of the posterior predictive fits and the error
 679 bars reflect the standard error. The risk-sensitive log relative value model explained the
 680 largest proportion ($R^2=0.85$) of the response gain variance of any model.

681 For comparison with experiments 1 and 2, we used ROC analysis to identify the time of first
 682 divergence in visuomotor responses between trials obtained in the blocks with most
 683 asymmetrical target values ($R_O/R_C = 0.5$, $P_O/P_C = 0.5$ versus $R_O/R_C = 2$, $P_O/P_C = 2$). The
 684 mean time of divergence was 151 ms for the asymmetrical value conditions, which was
 685 significantly later than the onset of the visuomotor response in a single target condition (120
 686 ms, paired t-test, $t_9 = -2.3$, $P = 0.045$).

687 The analysis of hand acceleration responses in standard trials was very similar to the force
 688 responses. Cross-condition regression analyses confirmed that the response size depended
 689 significantly on both probability and reward separately ($P < 10^{-5}$ & $P = 0.0008$ respectively)
 690 and that the risk-sensitive relative value model (Fig. 8D) explained the most variance in the
 691 acceleration data ($R^2=0.87$).

692 -- Insert Figure 8 about here --

693

694 Finally, we examined whether the ratio of hit-rates for outer to centre targets were related to
 695 the visuomotor response size across all reward and probability conditions (Figure 9). Linear
 696 regression showed that there was a significant linear relationship ($P = 0.002$) between
 697 visuomotor response size and the relative proportions of outer and centre target hits. Thus,
 698 rapid feedback response size was predictive of task success over a wide range of relative
 699 value conditions.

700 -- Insert Figure 9 about here --

701 **Comparisons across experiments: Task-dependent** 702 **response modulation**

703 There were differences in task features between the three experiments that comprise this
 704 study, as well as differences in absolute visuomotor response magnitudes (especially
 705 between experiments 1 & 2 versus experiment 3). It seems highly likely that the overall task
 706 context (e.g. how many potential targets were available, and the proportion of channel trials
 707 and standard trials with a cursor jump perturbation set for each target) was a key factor that
 708 mediated response magnitude across conditions and experiments - independently of value
 709 manipulations. For example, the visuomotor response magnitudes when the center target

was more rewarded or more probable in experiments 1 and 2 were smaller than those observed in the single target condition, despite a 20 fold larger reward, or 2 fold greater probability, for the centre target. By contrast, response magnitudes when the relative value for the centre target was greatest in experiment 3 were comparable to the single target condition. This general notion of task-dependent reflex modulation is consistent with a large body of published work on reflex function (eg Hammond, 1956), whereas our paper is focused on the influence of high and low value targets on visuomotor responses under *identical* task conditions in each experiment. The specific task constraints we used to interrogate the effects of value manipulations in each experiment almost certainly influenced the absolute response magnitudes. However, the influence of value on the relative response magnitudes was consistent across experiments despite the different task features. Therefore, the effects of value we describe are unlikely to depend strongly on the particular task configuration.

Discussion

Our results show that the size of rapid visuomotor responses to a sudden change in visual feedback about limb position depends upon the relative values of potential movement goals. Previous work showed that rapid feedback responses are flexibly regulated according to multiple components of the task, including timing (Franklin and Wolpert, 2008; Cluff and Scott, 2015) and spatial precision constraints (Gallivan et al., 2016), the presence of obstacles in the environment (Nashed et al., 2012, 2014), and the energetic or control costs associated with potential corrective responses (Nashed et al., 2012, 2014; Pruszynski et al., 2014). Rapid feedback responses can also reflect a control policy intermediate between those associated with competing goals when there is uncertainty about which goal will ultimately be specified (Gallivan et al., 2016). This illustrates that sensorimotor control policies are sufficiently flexible to take account of multiple potential goals simultaneously. However, this previous work which examined the flexibility of fast feedback responses manipulated physical characteristics of the task that are inherently coupled to the required motor outputs. Our current results show that feedback control is sensitive to decision variables, such as prospective reward, under otherwise identical task conditions. This suggests that feedback control policies that govern state-dependent transformations of sensory feedback to motor commands can be tailored to implement value-based choice.

Previous work showed that the magnitude rapid feedback responses to target perturbations or rapid target onsets was little affected by asymmetrical presentation probability (Reynolds and Day, 2012; Pruszynski et al., 2016). We see two potential explanations for this

discrepancy with our current findings about visuomotor responses to perturbations of hand position feedback. One possibility is that automatic responses to *target* jumps are insensitive to value asymmetry, in contrast to responses to our current results for *cursor* jumps. These different forms of perturbation appear to involve dissociable response processes ((Franklin et al., 2016), see (Oostwoud Wijdenes and Medendorp, 2017) for review). An alternative possibility, which we favour, is that automatic feedback responses are more strongly modulated on the basis of expected value when reflex tuning is necessary to achieve task success. Pilot experiments for this study suggested that the effects of reward on visuomotor responses were negligible when subjects could hit the target on every trial. This situation of assured success appears to be present in both the (Reynolds and Day, 2012; Pruszynski et al., 2016)et al (2016) and and the Reynolds and Day (2012) studies, neither of which enforced a time deadline for target acquisition. Nonetheless, it is important to note that our conclusions regarding the capacity for rapid feedback corrections to be modulated according to expected value may hold specifically for visual perturbations of limb position feedback under conditions of time pressure. More generally, while our laboratory task shows that value can affect reflexes under controlled conditions, the extent and importance of such modulation in real world tasks is still an open question.

Our results also extend previous findings that evolving decisions and reward or action history can bias the state of the motor system towards favoured potential actions during motor planning (Platt and Glimcher, 1999; Dorris et al., 2000; Lauwereyns et al., 2002; Milstein and Dorris, 2007; Marinovic et al., 2017). This previous work showed that neural activity in the saccadic and reach control networks is increased for neurons encoding high-value actions, and leads to biases in motor behaviour towards favoured actions. In our task, such neural activity cannot be biased in favour of a specific action (e.g. leftward or rightward hand force) as a function of reward or an ongoing decision, but rather decision variables must flexibly modulate the magnitude and direction of hand force as a function of evolving sensory input (i.e. depending on the direction of the cursor jump). The fast visuomotor response cannot therefore involve the relatively slow evaluation of evidence in higher-order brain areas traditionally associated with cognitive decision making. This does not imply that high-order cognitive areas are not critical for making value estimations, or even for setting reflex gains, prior to movement, to reflect trial-by-trial updates to value-based decision processes. Indeed, Franklin et al (2014) showed that the brain can learn different visuomotor response gains for left and right cursor perturbations depending on whether or not the perturbations in each direction were task relevant. This work showed that visuomotor responses can be modulated appropriately to hit targets when distinct perturbation types can be predicted based on the direction of initial cursor deviation. In the current study, we show that visuomotor responses

are modified for identical perturbations as a function of target value. Thus, it appears that low-level sensorimotor areas initiate motor decisions on the basis of the expected values of response alternatives, using the continuous stream of sensory evidence available during an individual movement. In this sense, fast visuomotor responses make a contribution to value based decision making.

If low-level sensorimotor circuits can contribute to value-based decisions through continuous feedback control, rather than merely executing the outcome of discrete action decisions taken in higher-order brain areas, it would support for the hypothesis that value-based decision algorithms are distributed throughout multiple levels of sensorimotor and cognitive processing hierarchies (Hunt et al., 2014; Hunt and Hayden, 2017). This notion differs from the traditional view that decisions arise from a serial process with modular units for choice evaluation, value comparison and action selection. According to the alternative view, the basis for decisions is mutual inhibition between neural representations of alternative options, and these computations occur simultaneously in multiple brain areas along both motor and abstract-value dimensions of tasks (Wang, 2012). Our current evidence that value-based decisions can be implemented through sensorimotor feedback control supports the alternative view, and the general notion that behaviour emerges via a distributed consensus between circuits engaged nominally in decision and sensorimotor processes (Cisek, 2012).

The regulation of rapid feedback responses was associated with successful pursuit of high-value task outcomes in our experiments, suggesting that value-based response tuning is functionally advantageous. This makes ethological sense in a dynamic world in which information about response options can change rapidly. For example, if a movement is perturbed such that an action associated with high reward becomes available, feedback corrections that oppose the perturbation should be inhibited in order to maximise the rewards obtained. This resembles a minimum intervention principle, in which errors are only corrected if they directly interfere with attainment of the task goal, and which is a hallmark of optimal feedback control systems (Todorov and Jordan, 2002; Diedrichsen, 2007; Liu and Todorov, 2007). In our task, when the cursor jumped to one side, the new lateral position of the cursor was often in a better state to maximise value than the unperturbed position, and so the jump was little corrected or even exaggerated if outer targets were much more valuable than the centre target. Thus, feedback control can not only reduce the costs associated with achieving a particular outcome, but at a broader level implement policies that include decision processes that evaluate and select goals (e.g.(Nagengast et al., 2010; Braun et al., 2011; Wolpert and Landy, 2012; Christopoulos and Schrater, 2015)).

819 A rational agent seeking to maximise cumulative rewards in the long run should make
 820 choices according to the relative value of available options. However, humans and other
 821 animals often behave according to risk-modulated value functions; they make choices that
 822 lead to lower overall gains by favouring larger, less certain rewards when risk-seeking, or
 823 more-certain but smaller rewards when risk-averse. Both risk-seeking and risk-averse
 824 behaviour have been exhibited in motor decision tasks, depending on factors such as the
 825 probability of successful outcomes (Trommershäuser et al., 2008; Wu et al., 2009;
 826 Nagengast et al., 2010, 2011b; McDougle et al., 2016). Interestingly, the valence of risk-
 827 modulation in motor tasks is often mirror opposite to that observed for economic decisions
 828 (Wu et al., 2009; McDougle et al., 2016). Our current results suggest that rapid feedback
 829 responses are tuned to a risk-averse value function. An interesting future question might be
 830 to determine whether an individual's risk sensitivity in visuomotor response regulation
 831 correlates with their risk sensitivity in cognitive decision making.

832
 833 An alternative perspective on the apparent risk-aversion evident for motor decisions in this
 834 study is that there is an “intrinsic” reward associated with successfully attaining a motor goal.
 835 Indeed, in Experiment 2, rapid visuomotor responses and task performance were biased
 836 toward more probable targets in the absence of any financial reward. If these results are to
 837 be interpreted in a value-based decision making framework, then a non-zero intrinsic reward
 838 component to value is obligatory. More broadly, the fact that humans sometimes decide to
 839 perform costly and difficult movements in the absence of explicit rewards, for example in
 840 (non-professional) sport and the performance arts, suggests a capacity to arbitrarily assign
 841 intrinsic value to completion of challenging physical tasks. Such flexibility in value
 842 assignment might be necessary to afford humans the capacity to make decisions based on
 843 complex reasoning or affect, than simply on the prospect of explicit reward. An important
 844 corollary of this idea is that if the attainment of intrinsic rewards is a general feature of
 845 successful completion of goal-directed sensorimotor tasks, experimental manipulation of
 846 reward through payment of money or provision of food would tend to underestimate the
 847 composite value derived from any given action. This scenario would complicate
 848 interpretations of choice behaviour involving motor tasks based on prospect theory.
 849 Nonetheless, our results clearly show that feedback control policies are biased to favour
 850 more frequent task success at the expense of the overall extrinsic reward accumulated
 851 through movement. This implies that low-level visuomotor feedback loops can reflect the
 852 outcomes of nuanced choice algorithms associated with value based decision making.

853

Figure Legends

Figure 1. Schematic of the task and conditions. A) Schematic of cursor and hand paths on a standard trial. Participants moved the cursor through a grey bar located 11 cm from the start position, at which point the cursor could jump either to the left (shown) or right or remain unperturbed. When the hand was 15.5 cm (or 16 cm - see methods) from the start, a yellow target was displayed in one of the three boxes according to probabilities shown in panel C. On cursor jump trials, participants had to move their hand in the opposite direction to the jump to acquire the centre target, but in the same direction to acquire an outer target (not shown). B) Schematic for a channel trial in which the hand was constrained to a straight line to the centre target by a force channel (blue arrows) generated by the robot. On cursor jump trials, the cursor was aligned to the final target when it was 3 cm from the target. On these trials we measured the lateral force produced against the walls of the channel. C) Manipulation of reward and target probabilities on unperturbed (left) and cursor jump (right) trials. In different experiments we manipulated the relative rewards (R_O/R_C) and target probabilities (P_O/P_C) associated with the outer and central targets. In the reward modulation experiment (Experiment 1) on unperturbed trials all target were possible with each target equally likely. On cursor jump trials, the center and outer target corresponding to the jump were equally likely. On all trials, the relative reward of the outer and central target had a twenty-fold difference. In the probability modulation experiment (Experiment 2) all targets had the same reward but the relative probability of an outer to central target had a two-fold difference. In the value modulation experiment (Experiment 3) both the reward and probability were manipulated on cursor jump trials (note that we did not include the condition in which relative reward and probability were both 1). In Experiment 3, on unperturbed trials only the center target was cued.

Figure 2. Force and acceleration responses to cursor jumps when different rewards were associated with centre and outer targets. Note that the centre target was ultimately cued on all trials shown here. Responses to left and right cursor jumps are pooled for all plots (right force signs are flipped), such that positive forces and accelerations represent a correction towards the center target. The grey highlighted zone indicates the period over which force and acceleration responses were averaged for analysis. A) Force response traces on every individual trial in which there was a cursor jump from an example subject. B) Average force responses across all trials for each participant - each trace is the average of all trials from centre rewarded, outer rewarded and single target conditions completed by a different participant. C) Group average \pm SEM lateral forces exerted into the wall of the force channel. Vertical dashed lines and shading illustrate the mean \pm SEM onset times of the visuomotor

889 response in the single target condition (grey) and the first time of response divergence
 890 between trial blocks with the outer versus centre targets more probable (green). D) Group
 891 averages (bars) and individual subject averages (grey lines) of the mean force exerted
 892 against the channel wall between 170 and 220 ms after the cursor jump. E) Group average \pm
 893 SEM hand acceleration traces on standard trials, when participants had to correct for the
 894 cursor jump in order to hit the target, as in figure 1A. Vertical lines depict response onsets as
 895 per panel D. F) Corresponding hand acceleration group averages (bars) and individual
 896 subject averages (grey lines).

897
 898 **Figure 3.** Force and acceleration responses to cursor jumps when centre and outer targets
 899 had different probabilities of selection. Sign conventions are as for Figure 2. A) Force
 900 response traces on every individual trial in which there was a cursor jump from an example
 901 subject. B) Average force responses across all trials for each participant - each trace is the
 902 average of all trials from centre more probable, outer more probable and single target
 903 conditions completed by a different participant. C) Group average \pm SEM lateral forces
 904 exerted into the wall of the force channel. Vertical dashed lines and shading illustrate the
 905 mean \pm SEM onset time of the visuomotor response in the single target condition (grey) and
 906 the first time of response divergence between trial blocks with the outer versus centre
 907 targets more probable (green) D) Group averages (bars) and individual subject averages
 908 (grey lines) of the mean force exerted against the channel wall between 170 and 220 ms
 909 after the cursor jump. E) Group average \pm SEM hand acceleration traces on standard trials,
 910 when participants had to correct for the cursor jump in order to hit the target. Vertical lines
 911 and shading as per panel D. F) Corresponding and acceleration group averages (bars) and
 912 individual subject averages (grey lines).

913 **Figure 4.** Force responses to left and right cursor jumps in Experiment 3 when the centre
 914 and outer targets were ultimately specified. Group average \pm SEM lateral forces exerted into
 915 the wall of the force channel show that the analysis window (170 - 220 ms after the cursor
 916 jump) was prior to the first point at which the force responses for centre and outer targets
 917 diverged (red dotted line and shaded region represent the mean \pm SEM of the point of first
 918 divergence as assessed by ROC analysis).

919 **Figure 5.** Force responses to cursor jumps for all eight conditions in which the probabilities
 920 and rewards associated with centre and outer targets were varied. A. Each panel shows
 921 data for trials with different reward ratios, with traces within a panel grouped by different
 922 target probability ratios. These show group averages \pm SEM for the lateral forces exerted

923 into the wall of the force channel. B. Each panel shows data for trials with different target
924 probability ratios, with traces within a panel grouped by different reward ratios.

925 **Figure 6.** Acceleration responses to cursor jumps for all eight conditions in which the
926 probabilities and rewards associated with centre and outer targets were varied. A. Each
927 panel shows data for trials with different reward ratios, with traces within a panel grouped by
928 different target probability ratios. These show group averages \pm SEM for the lateral
929 acceleration. B. Each panel shows data for trials with different target probability ratios, with
930 traces within a panel grouped by different reward ratios.

931

932 **Figure 7.** Visuomotor force responses in channel trials to varying both reward and
933 probability ratios in Experiment 3. Response size against (A) log relative probability, (B) log
934 relative reward, (C) log relative value, and (D) log risk-modulated relative value. 50% and
935 95% bootstrapped confidence regions are shaded in grey (darker, and lighter regions
936 respectively).

937 **Figure 8.** Visuomotor acceleration responses in standard trials to varying both reward and
938 probability ratios in Experiment 3. Response size against (A) log relative probability, (B) log
939 relative reward, (C) log relative value, and (D) log risk-modulated relative value. 50% and
940 95% bootstrapped confidence regions are shaded in grey (darker, and lighter regions
941 respectively).

942 **Figure 9.** Target hit rate depends on the visuomotor response magnitude. The group
943 average (\pm SEM) hit rate ratios are defined as the hit rate for outer targets (t_o) divided by the
944 hit rate for centre targets (t_c). These are plotted against the visuomotor response size (\pm
945 SEM) for all combinations of relative reward and probability in Experiment 3. More negative
946 visuomotor responses reflect greater correction of the cursor jump toward the centre target.
947 The line shows the linear regression fit.

948

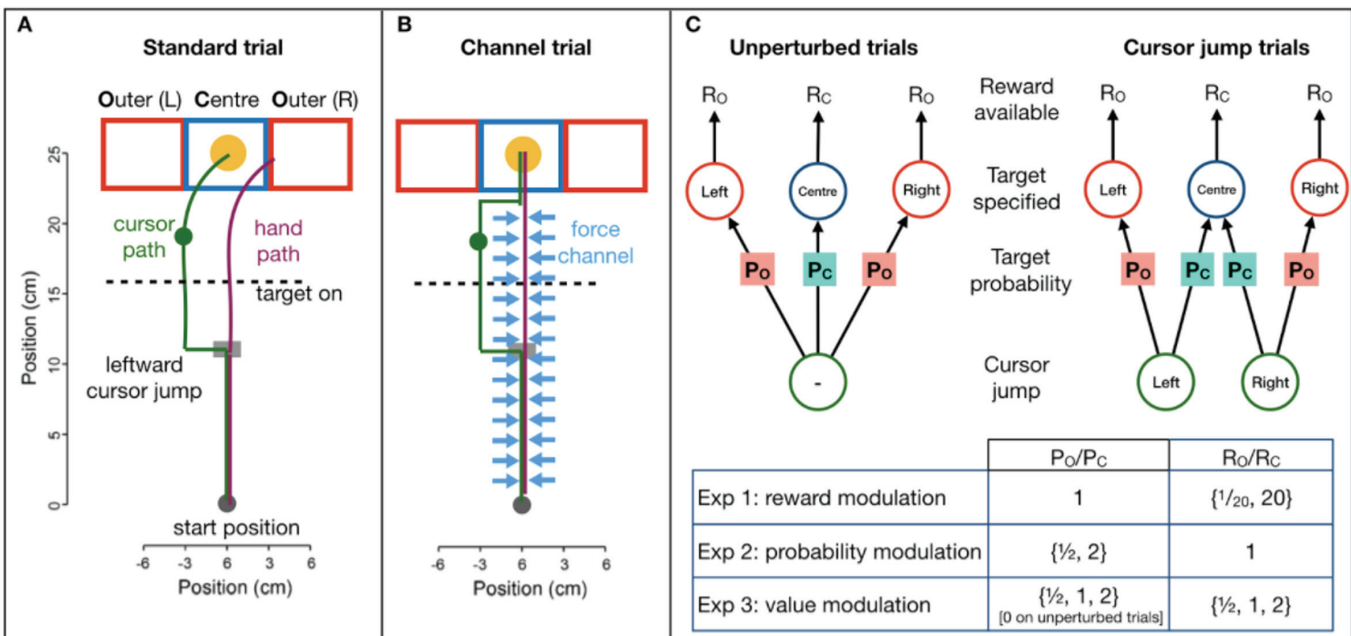
949 **References**

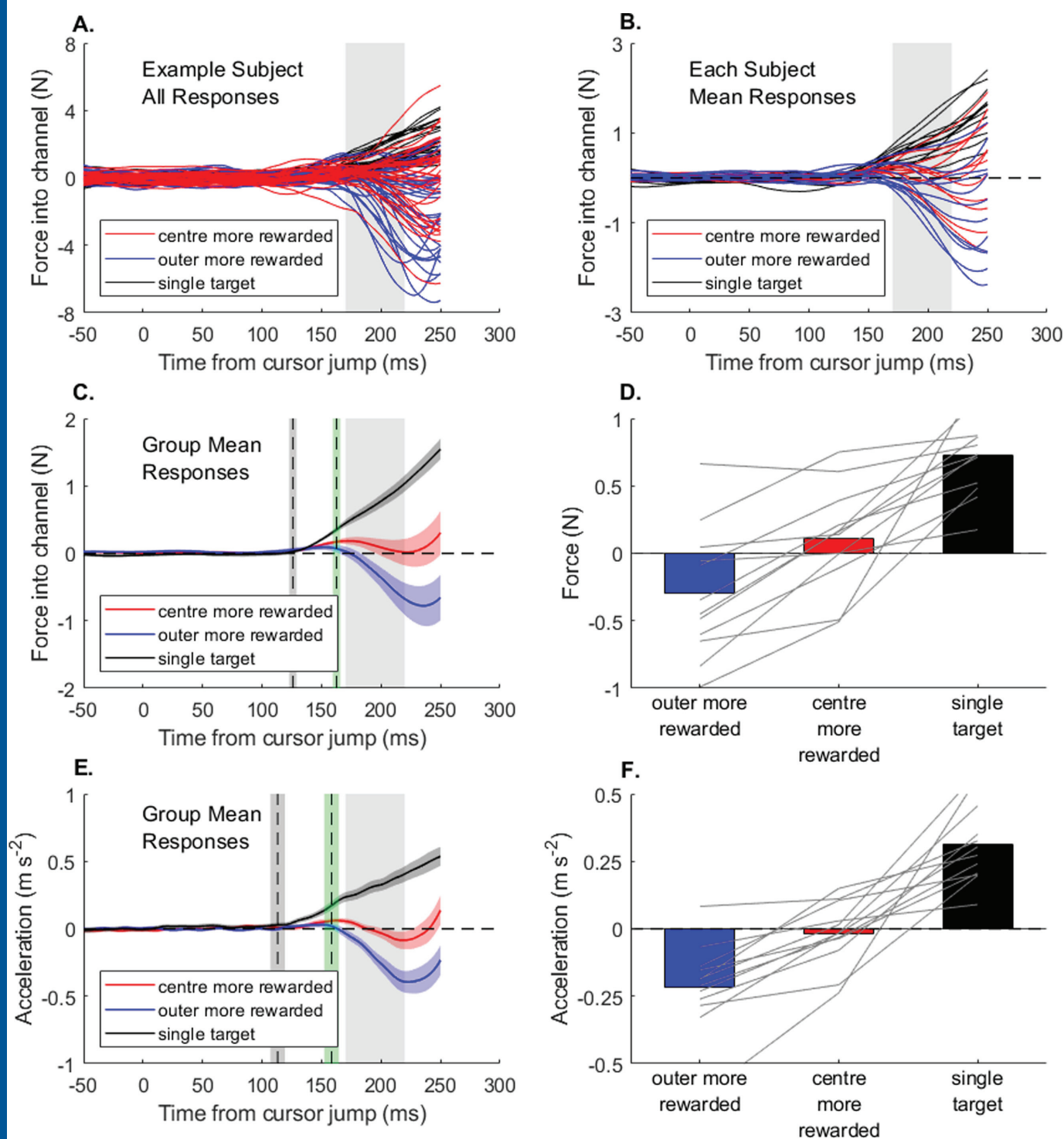
- 950 Braun DA, Nagengast AJ, Wolpert DM (2011) Risk-sensitivity in sensorimotor control. *Front*
951 *Hum Neurosci* 5:1.
- 952 Christopoulos V, Schrater PR (2015) Dynamic Integration of Value Information into a
953 Common Probability Currency as a Theory for Flexible Decision Making. *PLoS Comput*
954 *Biol* 11:e1004402.
- 955 Cisek P (2012) Making decisions through a distributed consensus. *Curr Opin Neurobiol*
956 22:927–936.
- 957 Cluff T, Scott SH (2015) Apparent and Actual Trajectory Control Depend on the Behavioral
958 Context in Upper Limb Motor Tasks. *J Neurosci* 35:12465–12476.
- 959 Croxson PL, Walton ME, O'Reilly JX, Behrens TEJ, Rushworth MFS (2009) Effort-based
960 cost-benefit valuation and the human brain. *J Neurosci* 29:4531–4541.
- 961 Day BL, Lyon IN (2000) Voluntary modification of automatic arm movements evoked by
962 motion of a visual target. *Exp Brain Res* 130:159–168.
- 963 Diedrichsen J (2007) Optimal task-dependent changes of bimanual feedback control and
964 adaptation. *Curr Biol* 17:1675–1679.
- 965 Dimitriou M, Wolpert DM, Franklin DW (2013) The temporal evolution of feedback gains
966 rapidly update to task demands. *J Neurosci* 33:10898–10909.
- 967 Dorris MC, Paré M, Munoz DP (2000) Immediate neural plasticity shapes motor
968 performance. *J Neurosci* 20:RC52.
- 969 Franklin DW, Franklin S, Wolpert DM (2014) Fractionation of the visuomotor feedback
970 response to directions of movement and perturbation. *J Neurophysiol* 112:2218–2233.
- 971 Franklin DW, Reichenbach A, Franklin S, Diedrichsen J (2016) Temporal Evolution of Spatial
972 Computations for Visuomotor Control. *Journal of Neuroscience* 36:2329–2341.
- 973 Franklin DW, Wolpert DM (2008) Specificity of reflex adaptation for task-relevant variability. *J*
974 *Neurosci* 28:14165–14175.
- 975 Gallivan JP, Logan L, Wolpert DM, Flanagan JR (2016) Parallel specification of competing
976 sensorimotor control policies for alternative action options. *Nat Neurosci* 19:320–326.
- 977 Hammond PH (1956) The influence of prior instruction to the subject on an apparently
978 involuntary neuro-muscular response. *J Physiol* 132:17–18P.
- 979 Howard IS, Ingram JN, Wolpert DM (2009) A modular planar robotic manipulandum with
980 end-point torque control. *J Neurosci Methods* 181:199–211.
- 981 Hunt LT, Dolan RJ, Behrens TEJ (2014) Hierarchical competitions subserving multi-attribute
982 choice. *Nat Neurosci* 17:1613–1622.
- 983 Hunt LT, Hayden BY (2017) A distributed, hierarchical and recurrent framework for reward-
984 based choice. *Nat Rev Neurosci* 18:172–182.
- 985 Kahneman D (2003) Maps of bounded rationality: psychology for behavioral economics. *Am*
986 *Econ Rev* 93:1449–1475.

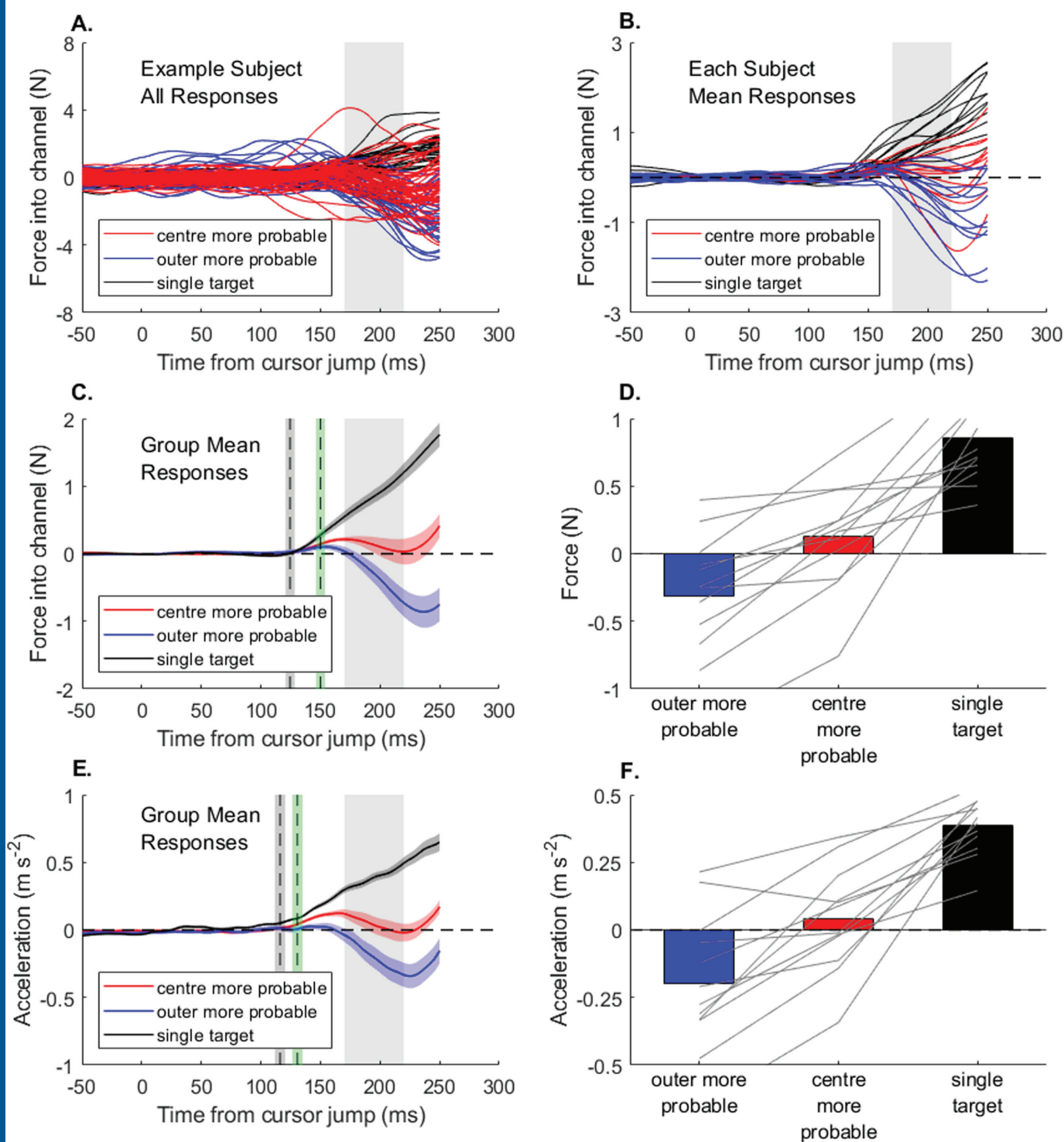
- 987 Kass RE, Raftery AE (1995) Bayes Factors. *J Am Stat Assoc* 90:773–795.
- 988 Klein-Flügge MC, Kennerley SW, Friston K, Bestmann S (2016) Neural Signatures of Value
989 Comparison in Human Cingulate Cortex during Decisions Requiring an Effort-Reward
990 Trade-off. *J Neurosci* 36:10002–10015.
- 991 Kruschke JK (2010) *Doing Bayesian Data Analysis*.
- 992 Landy MS, Trommershäuser J, Daw ND (2012) Dynamic estimation of task-relevant
993 variance in movement under risk. *J Neurosci* 32:12702–12711.
- 994 Lauwereyns J, Watanabe K, Coe B, Hikosaka O (2002) A neural correlate of response bias
995 in monkey caudate nucleus. *Nature* 418:413–417.
- 996 Liu D, Todorov E (2007) Evidence for the flexible sensorimotor strategies predicted by
997 optimal feedback control. *J Neurosci* 27:9354–9368.
- 998 Manohar SG, Chong TT-J, Apps MAJ, Batla A, Stamelou M, Jarman PR, Bhatia KP, Husain
999 M (2015) Reward Pays the Cost of Noise Reduction in Motor and Cognitive Control.
1000 *Curr Biol* 25:1707–1716.
- 1001 Marinovic W, Poh E, de Rugy A, Carroll TJ (2017) Action history influences subsequent
1002 movement via two distinct processes. *Elife* 6 Available at:
1003 <http://dx.doi.org/10.7554/eLife.26713>.
- 1004 McDougale SD, Boggess MJ, Crossley MJ, Parvin D, Ivry RB, Taylor JA (2016) Credit
1005 assignment in movement-dependent reinforcement learning. *Proc Natl Acad Sci U S A*
1006 113:6797–6802.
- 1007 Milstein DM, Dorris MC (2007) The influence of expected value on saccadic preparation. *J*
1008 *Neurosci* 27:4810–4818.
- 1009 Nagengast AJ, Braun DA, Wolpert DM (2010) Risk-sensitive optimal feedback control
1010 accounts for sensorimotor behavior under uncertainty. *PLoS Comput Biol* 6:e1000857.
- 1011 Nagengast AJ, Braun DA, Wolpert DM (2011a) Risk-sensitivity and the mean-variance trade-
1012 off: decision making in sensorimotor control. *Proc Biol Sci* 278:2325–2332.
- 1013 Nagengast AJ, Braun DA, Wolpert DM (2011b) Risk sensitivity in a motor task with speed-
1014 accuracy trade-off. *J Neurophysiol* 105:2668–2674.
- 1015 Nashed JY, Crevecoeur F, Scott SH (2012) Influence of the behavioral goal and
1016 environmental obstacles on rapid feedback responses. *J Neurophysiol* 108:999–1009.
- 1017 Nashed JY, Crevecoeur F, Scott SH (2014) Rapid online selection between multiple motor
1018 plans. *J Neurosci* 34:1769–1780.
- 1019 Oostwoud Wijdenes L, Medendorp WP (2017) State Estimation for Early Feedback
1020 Responses in Reaching: Intramodal or Multimodal? *Front Integr Neurosci* 11:38.
- 1021 Pastor-Bernier A, Cisek P (2011) Neural correlates of biased competition in premotor cortex.
1022 *J Neurosci* 31:7083–7088.
- 1023 Platt ML, Glimcher PW (1999) Neural correlates of decision variables in parietal cortex.
1024 *Nature* 400:233–238.
- 1025 Pruszyński JA, Johansson RS, Flanagan JR (2016) A Rapid Tactile-Motor Reflex

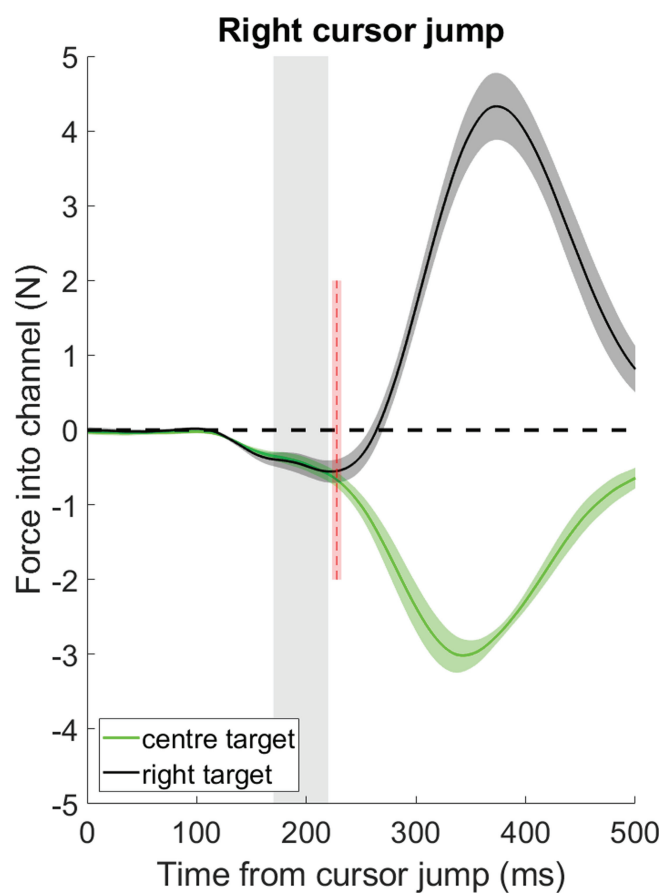
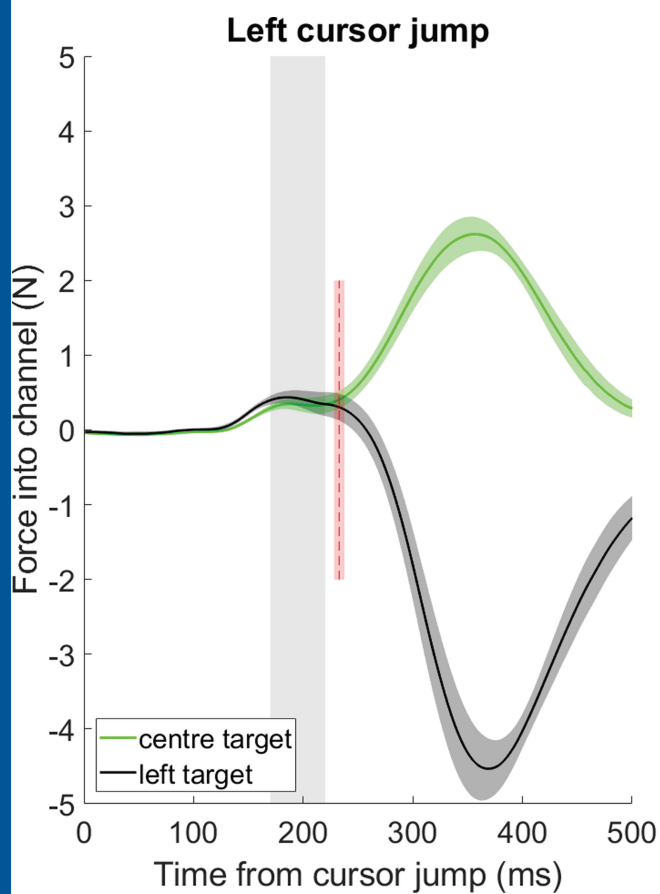
- 1026 Automatically Guides Reaching toward Handheld Objects. *Curr Biol* 26:788–792.
- 1027 Pruszynski JA, Omrani M, Scott SH (2014) Goal-dependent modulation of fast feedback
1028 responses in primary motor cortex. *J Neurosci* 34:4608–4617.
- 1029 Resulaj A, Kiani R, Wolpert DM, Shadlen MN (2009) Changes of mind in decision-making.
1030 *Nature* 461:263–266.
- 1031 Reynolds RF, Day BL (2012) Direct visuomotor mapping for fast visually-evoked arm
1032 movements. *Neuropsychologia* 50:3169–3173.
- 1033 Salvatier J, Wiecki TV, Fonnesbeck C (2016) Probabilistic programming in Python using
1034 PyMC3. *PeerJ Computer Science* 2:e55.
- 1035 Schütz AC, Trommershäuser J, Gegenfurtner KR (2012) Dynamic integration of information
1036 about salience and value for saccadic eye movements. *Proc Natl Acad Sci U S A*
1037 109:7547–7552.
- 1038 Scott SH (2004) Optimal feedback control and the neural basis of volitional motor control.
1039 *Nat Rev Neurosci* 5:532–546.
- 1040 Scott SH (2016) A Functional Taxonomy of Bottom-Up Sensory Feedback Processing for
1041 Motor Actions. *Trends Neurosci* 39:512–526.
- 1042 Selen LPJ, Shadlen MN, Wolpert DM (2012) Deliberation in the motor system: reflex gains
1043 track evolving evidence leading to a decision. *J Neurosci* 32:2276–2286.
- 1044 Seydell A, McCann BC, Trommershäuser J, Knill DC (2008) Learning stochastic reward
1045 distributions in a speeded pointing task. *J Neurosci* 28:4356–4367.
- 1046 Shadmehr R, Huang HJ, Ahmed AA (2016) A Representation of Effort in Decision-Making
1047 and Motor Control. *Curr Biol* 26:1929–1934.
- 1048 Skvortsova V, Palminteri S, Pessiglione M (2014) Learning to minimize efforts versus
1049 maximizing rewards: computational principles and neural correlates. *J Neurosci*
1050 34:15621–15630.
- 1051 Song J-H, Nakayama K (2008) Target selection in visual search as revealed by movement
1052 trajectories. *Vision Res* 48:853–861.
- 1053 Todorov E (2004) Optimality principles in sensorimotor control. *Nat Neurosci* 7:907–915.
- 1054 Todorov E, Jordan MI (2002) Optimal feedback control as a theory of motor coordination.
1055 *Nat Neurosci* 5:1226–1235.
- 1056 Trommershäuser J, Maloney LT, Landy MS (2008) Decision making, movement planning
1057 and statistical decision theory. *Trends Cogn Sci* 12:291–297.
- 1058 Vehtari A, Gelman A, Gabry J (2017) Practical Bayesian model evaluation using leave-one-
1059 out cross-validation and WAIC. *Stat Comput* 27:1413–1432.
- 1060 Wang X-J (2012) Neural dynamics and circuit mechanisms of decision-making. *Curr Opin*
1061 *Neurobiol* 22:1039–1046.
- 1062 Weiler J, Gribble PL, Pruszynski JA (2015) Goal-dependent modulation of the long-latency
1063 stretch response at the shoulder, elbow, and wrist. *J Neurophysiol* 114:3242–3254.

- 1064 Wolpert DM, Landy MS (2012) Motor control is decision-making. *Curr Opin Neurobiol*
1065 22:996–1003.
- 1066 Wu S-W, Delgado MR, Maloney LT (2009) Economic decision-making compared with an
1067 equivalent motor task. *Proc Natl Acad Sci U S A* 106:6088–6093.
- 1068 Yang T, Shadlen MN (2007) Probabilistic reasoning by neurons. *Nature* 447:1075–1080.
1069

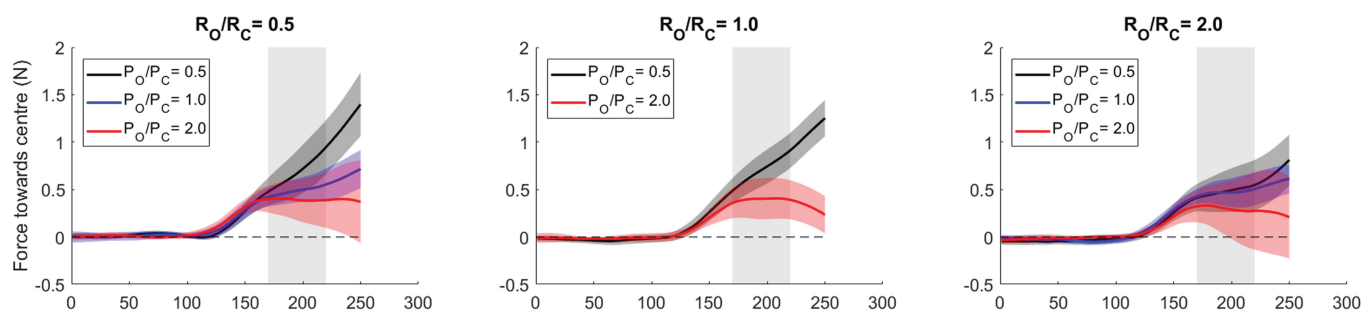




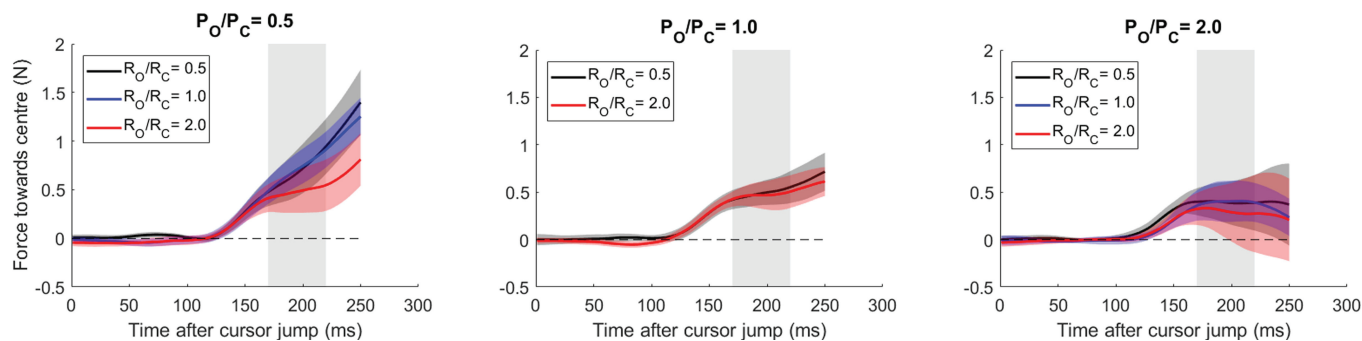




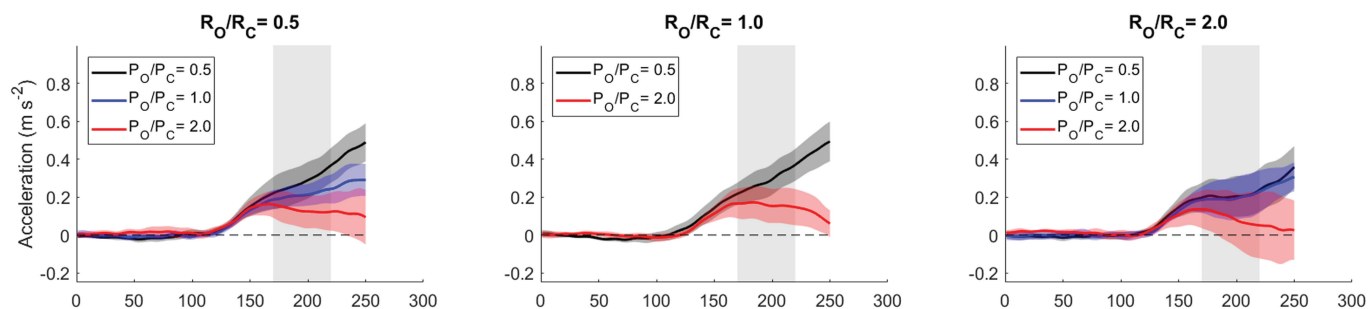
A. Fixed Reward Ratios



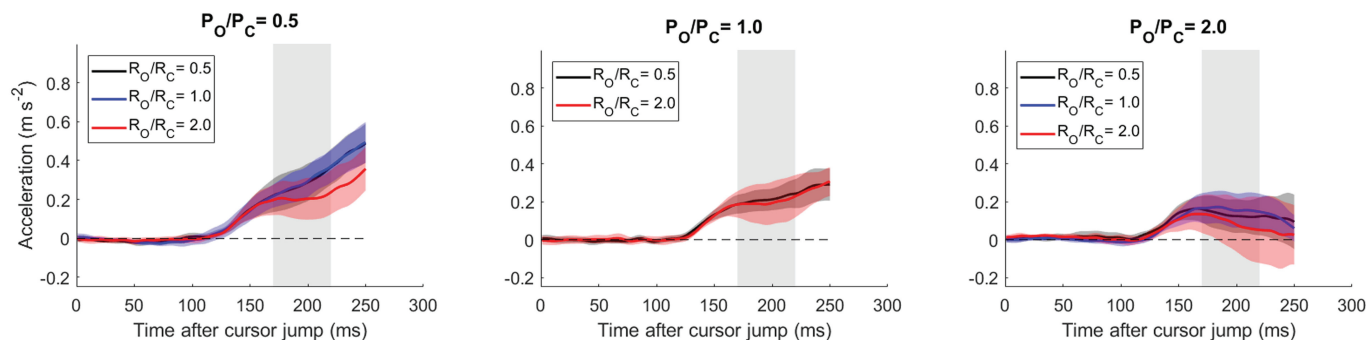
B. Fixed Probability Ratios

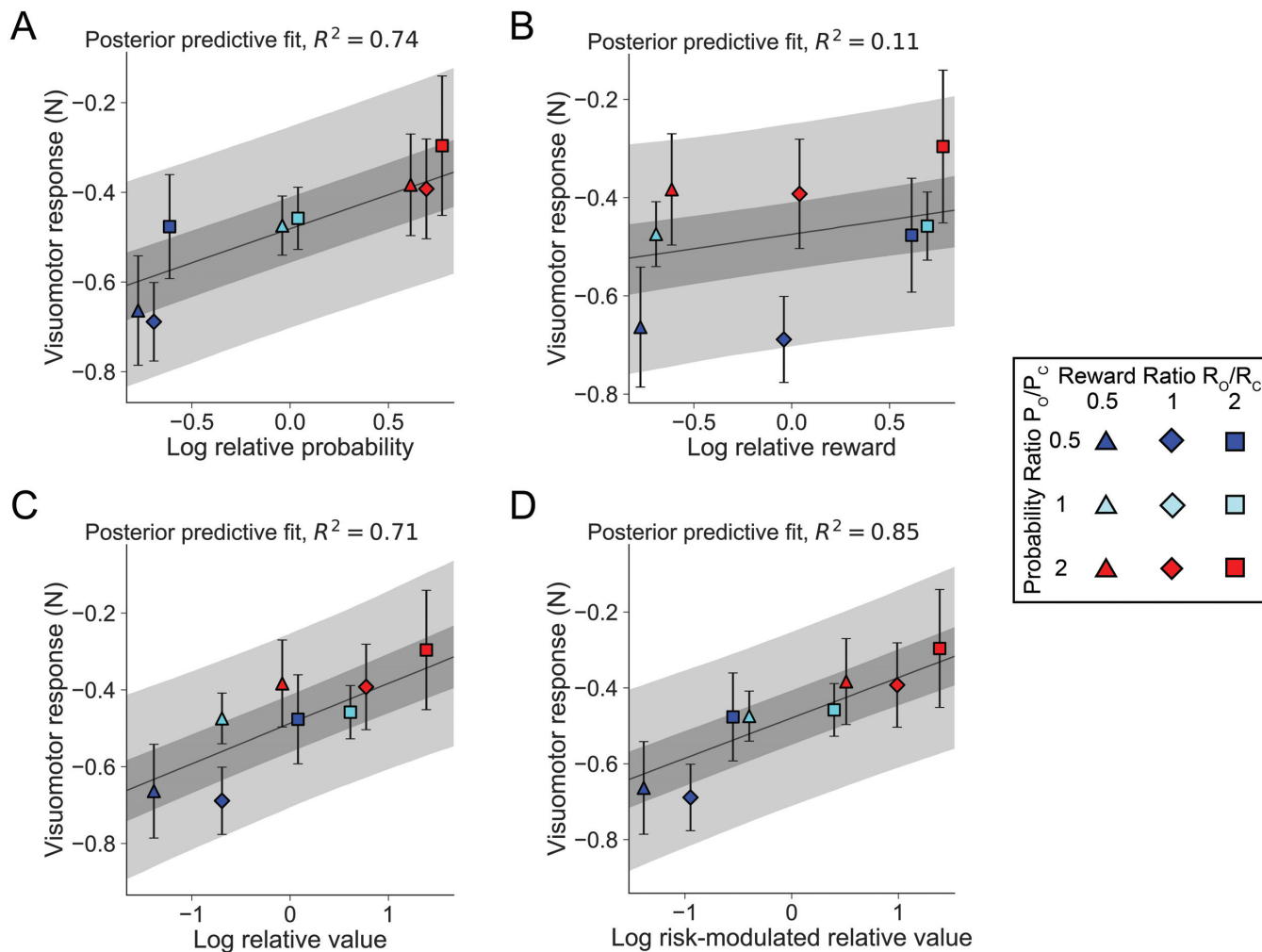


A. Fixed Reward Ratios

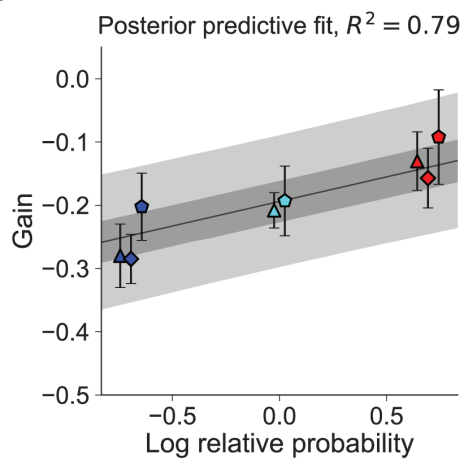


B. Fixed Probability Ratios

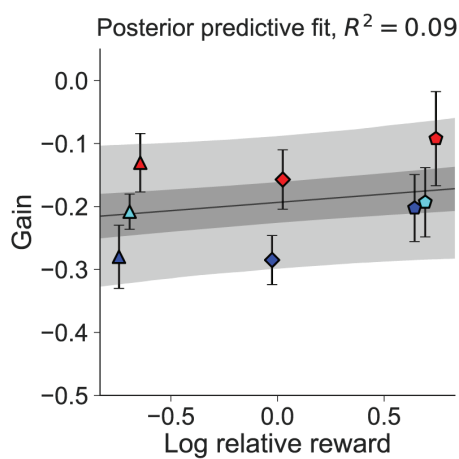




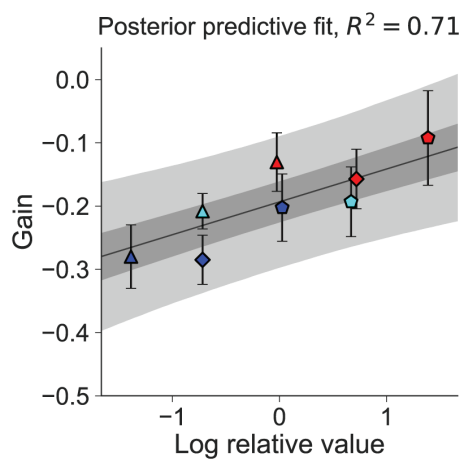
A



B



C



D

